Harmonic Modes as Variables to Approximately Account for Receptor Flexibility in Ligand–Receptor Docking Simulations: Application to DNA Minor Groove Ligand Complex

MARTIN ZACHARIAS, HEINZ SKLENAR

Max Delbrück Center for Molecular Medicine, Theoretical Biophysics Group, Robert Rössle Strasse 10, D-13122 Berlin, Germany

Received 15 October 1997; accepted 15 August 1998

ABSTRACT: An approach to approximately account for receptor flexibility in ligand–receptor docking simulations is described and applied to a DNA/Hoechst 33258 analogue complex. Harmonic modes corresponding to eigenvectors with small eigenvalues of the Hessian matrix of the potential energy function were used as independent variables to describe receptor flexibility. For the DNA minor groove ligand case most of the conformational difference between an energy minimized free DNA and ligand-bound structure could be assigned to 5–40 harmonic receptor modes with small eigenvalues. During docking, deformations of the DNA receptor structure in the subset of harmonic modes were limited using a simple penalty function that avoided the summation over all intrareceptor atom pairs. Significant improvement of the sterical fit between ligand and receptor was found upon relaxation of the DNA in the subset of harmonic modes after docking of the ligand at the position found in the known crystal structure. In addition, the harmonic mode relaxation resulted in DNA structures that were more similar to the energy minimized ligand-bound form. Although harmonic mode relaxation also leads to improved sterical fit for other ligand placements, the placement as observed in the crystal structure could still be identified as the site with the most favorable sterical interactions. Because relaxation in the harmonic modes is orders of magnitude faster than conventional energy minimization using all atom coordinates as independent variables, the approach might be useful as a preselection tool to

Correspondence to: M. Zacharias
Contract/grant sponsor: Deutsche Forschungsgemeinschaft

CCC 0192-8651 / 99 / 030287-14

recognize ligand binding sites accessible only upon small conformational changes of the receptor. The harmonic mode relaxed structures can only be considered as approximate structures because deformation of the receptor in the harmonic modes can lead to small perturbations of the stereochemical geometry of the molecule. Energy minimization of preselected ligand–DNA docking candidates in all atom coordinates is required to reduce these deviations. © 1999 John Wiley & Sons, Inc. J Comput Chem 20: 287–300, 1999

Keywords: ligand–receptor docking; normal mode analysis; DNA–ligand interaction; drug design; receptor flexibility

Introduction

The prediction of possible ligand binding sites on a receptor molecule such as DNA or protein is of great importance for the selection and design of molecules interfering with its biological function.[1–6] The ligand–receptor binding strength is determined by various energetic contributions, for example, Coulomb interactions between charges on the receptor and ligand, electrostatic and nonelectrostatic (hydrophobic) solvation terms, and entropic contributions.[3] Atomic resolution structural studies using X-ray crystallography or NMR of several receptor–ligand complexes show that steric complementarity is an important prerequisite for ligand binding. In addition, receptor as well as ligand can undergo conformational changes upon complex formation. For many cases of known X-ray or NMR receptor structures in the ligand-free and ligand-bound forms the observed ligand induced conformational changes are small (< 1 Å). However, these conformational changes are often nonlocal and involve small rearrangements of structural domains. In most current computational ligand–receptor docking procedures the receptor, as well as the ligand or ligand fragment, are treated as rigid bodies and possible binding sites are selected from a systematic search of sterically allowed docking sites on the receptor.[1,2,6] The receptor structure can either be the NMR or X-ray structure of the biomolecule or an energy minimized model built structure. To allow for some ligand flexibility the ligand molecule can be separated into fragments for which the assumption of a rigid structure is reasonable. The receptor structure is searched for binding sites of these fragments that in a subsequent step are linked together, if possible, through small linker fragments.[7–10] This procedure can overlook binding sites that are sterically accessible only upon small conformational rearrangements of the receptor. In addition, putative binding cavities assigned a low binding score due to a small number of receptor–ligand contacts might become very promising sites if some structural changes in the receptor would be allowed. Only a few docking studies have been published so far that account for some flexibility of receptor and/or ligand close to the binding region.[11–14] A particular noteworthy approach is the method by Knegtel et al.[14] Here the receptor target is not only represented by one experimentally determined structure but also by an ensemble of structures. The ensemble of structures either fulfills NMR spectroscopic data or is formed by combining X-ray structures solved under various crystallization conditions. This method is similar in speed to rigid docking methods but requires the availability of an ensemble of structures that appropriately represents the conformational flexibility of the receptor.

In the case of only one available receptor structure, treating all receptor atom coordinates explicitly during the docking process is computationally very demanding. Not only intermolecular interactions between receptor and ligand but also intramolecular interactions between receptor atoms need to be calculated many times. Because the number of interacting atom pairs grows with the square of the number of atoms this becomes rapidly computationally prohibitive for larger receptor molecules and many possible ligands and binding positions. A less computationally demanding approach is desirable to account at least approximately for conformational changes in the receptor molecule during docking simulations.

For a receptor structure at an energy minimum for a given force field, the possible directions of largest mobility of the molecule can be extracted from the second derivative matrix of the potential function (Hessian) through matrix diagonalization. Here, directions of collective motions are identified as eigenvectors of the Hessian matrix with

eigenvalues related to the expected energy change upon deformation of the structure in the corresponding eigenvector direction. Including an atom mass matrix, the approach, termed normal mode analysis, has been used to characterize collective motions in proteins[15–18] and nucleic acids.[19,20] Related techniques to extract directions of largest mobility from molecular dynamics simulations have been termed quasiharmonic, principal component, and essential dynamics methods to analyze simulations.[21–25] Here the variance–covariance matrix of atomic fluctuations is analyzed in the same way as the Hessian matrix of the potential function in the harmonic normal mode analysis. The technique is called quasiharmonic analysis because the motion in the quasiharmonic modes can be strongly anharmonic, but the statistical analysis is simply limited to the second moment [(co)variance] of the conformational fluctuations during the molecular dynamics simulation.[23–25]

An important result of these studies is that a large part of the motions of proteins and DNA (in the vicinity of an energy minimum or equilibrium structure) can be characterized by only a few collective modes. This observation leads to the idea to limit possible conformational changes in the receptor upon ligand binding to a subspace of collective modes corresponding to a set of eigenvectors of the Hessian matrix with small eigenvalues. In addition, treating these modes as independent variables for small moves using a simple penalty function to limit conformational changes in these modes avoids the summation over all intrareceptor atom pairs to calculate the receptor energy. The most time consuming parts of the approach are the minimization of the (uncomplexed) receptor structure, evaluation of the Hessian matrix, and matrix diagonalization; they can be done once in advance of the docking procedure. The same set of eigenvectors (harmonic modes, HMs) to describe conformational changes of the receptor can be used for all subsequent ligand–receptor docking attempts for various possible ligands and ligand starting positions. In a second step selected ligand–receptor complexes obtained after relaxation in the set of eigenvectors can be analyzed more accurately using, for example, energy minimization including pairwise intrareceptor interactions. In the present study the idea of using HMs as variables in a receptor–ligand docking approach was tested for the binding of a ligand, a Hoechst 33258 analogue,[26] to DNA as an example. This bis-benzimidazole compound differs from the original Hoechst 33258 in an imidazoline group replacing the piperazine ring in Hoechst 33258[26]. It belongs to a class of compounds that bind preferentially to AT regions at the minor groove of DNA. The structure of DNA in complex with minor groove ligands was studied extensively using X-ray crystallography[27–32] and NMR spectroscopy.[33–35] (See ref. 36 for a review on minor groove DNA recognition.) Comparison of crystal structures of free DNA and Hoechst 33258 analogue bound to DNA indicates that the DNA undergoes small but significant conformational changes upon minor groove ligand binding.[26,36] A second reason for choosing this system as a test case is that a DNA receptor fragment containing the ligand binding site is small enough to allow a relatively rapid calculation of the HMs and to study the effect of force field modifications on the approach.

Methods

The crystal structure of the Hoechst 33258 minor groove analogue {5-(-2-Imidazolinyl-2[2-(hydroxyphenyl)-5-benzimidazoly]benzimidazole} in complex with a dodecamer double stranded oligonucleotide $d(CGCGAATTCGCG)_2$[26] served as a reference structure. The ligand contacts the central 4 base pairs in the crystal structure. Energy minimization down to very small residual potential energy gradients and evaluation of the Hessian matrix and matrix diagonalization can become computationally demanding for systems with a large number of atoms. Therefore, the DNA used as a model in the present study was limited to the central 6 base pairs [$d(GAATTC)_1$]. A canonical B-DNA conformation of the oligonucleotide was generated using the molecular modeling program QUANTA.[37] The GROMOS87 force field[38] was used for energy minimization with the following modifications. Because no explicit solvent molecules or ions were included in the present study, solvent and salt screening effects were taken into account implicitly by decreasing the partial charges on the atoms (corresponding to an increase of the dielectric constant). To minimize the end effects of the relatively short piece of DNA and to prevent dissociation of the Watson–Crick type base pairing due to the charge scaling, a number of distance constraints were introduced. A quadratic distance constraining potential was added to all Watson–Crick type hydrogen bonds of the form: $V = k\ (d_{\text{H-A}} - d_{\text{o}})^2$. A distance $d_{\text{o}} = 1.8$ Å was

used as the optimal hydrogen bonding distance ($k = 100$ kcal mol^{-1} Å^{-2}). A similar quadratic distance constraining potential was used to constrain the distance between C1′ atoms in nucleotides of opposite strands ($d_o = 11.0$ Å, $k =$ 50 kcal mol^{-1} Å^{-2}). Addition of this penalty keeps the base pairs in an approximately planar geometry that mimics to a certain degree the effect of a continues helical structure (reducing end effects). To check the sensitivity of the approach to variations in the potential function and minimum energy structure, four force field modifications (A–D) were considered. Test calculations on the DNA fragment showed that a scaling of the partial atomic charges to mimic the dielectric damping by the surrounding solvent of > 0.35 resulted in conformations with a slightly opened minor groove. A charge scaling by a factor of < 0.35 yielded structures with a smaller minor groove compared to regular B-DNA (see also Results). In the force field model A the charges were scaled by a factor of 0.37. In models B–D the charges were scaled down to 0.33. For models C and D the improper torsion and angle force constants were increased by a factor of 3 and 4, respectively.

Energy minimization was performed using a conjugated gradient or variable metric method[39] down to gradients of $< 10^{-4}$ kcal mol^{-1} Å^{-1}. The Hessian matrix was evaluated analytically with appropriate expressions for the bonded and nonbonded terms in the GROMOS force field using self-written code. The matrix was diagonalized using standard methods.[39] Six eigenvectors with eigenvalues close to zero (at least 3 orders of magnitude smaller than eigenvalues of other eigenvectors) were identified as eigenvectors corresponding to overall translation and rotation of the molecule. A subset of eigenvectors was used as additional variables in a docking approach of the minor groove ligand and DNA. Note that in a standard normal mode analysis the moment of inertia tensor of the molecule is accounted for to calculate vibrational (normal mode) frequencies from the eigenvalues. This step was omitted because the focus of the current study was not the (vacuum) dynamics of the receptor molecule. Instead, the eigenvectors of the Hessian matrix were used in the present study and were termed HMs.

The minor groove ligand was kept rigid in its conformation found in the X-ray structure in all docking and energy minimization attempts. Similar to most simple docking procedures, only sterical interactions between the DNA and ligand were considered in the form of a Lennard–Jones potential (V_{LJ}) using the Lennard–Jones parameters $A = 2906.0$ kcal$^{0.5}$ mol$^{-0.5}$ Å^6, $B = 46.6$ kcal$^{0.5}$ mol$^{-0.5}$ Å^3 for all ligand atoms [with the $V_{LJ}(i, j) = A_i A_j r_{ij}^{-12} - B_i B_j r_{ij}^{-6}$, r_{ij} is the distance between atoms i and j].

A simple standard docking attempt consisted of two steps. The ligand was first placed at a certain position and orientation relative to the DNA that usually corresponded to the position and orientation found in the X-ray structure of the complex. In addition, the ligand position could also be energy optimized with respect to translational and orientational degrees of freedom relative to the receptor structure. In a second step the DNA receptor structure was subject to relaxation in the subset of HMs used to approximately describe the flexibility of the receptor. The force in the HM direction has an intrinsic component that is the derivative (gradient) of the penalty potential for deviations of the minimum structure in the HM direction and an external component due to the interaction of receptor atoms with ligand atoms. The force was calculated as the projection (scalar product) onto the corresponding HM direction. The harmonic approximation for the intramolecular energy change of the receptor upon deformation in a given HM suggests a quadratic penalty function with the force constant given by the corresponding eigenvalue. Test calculations showed that a more useful penalty function is of order 4 dependence for deviations in the eigenvector direction and a scaling of the force constant (eigenvalue) by 0.05. This penalty function allows for small conformational changes (a few tenths of an angstrom) in a set of HMs with a small increase of the penalty energy and a steep rise in energy for large conformational changes. Relaxation in the HMs was performed using conjugated gradient energy minimization in the HM variables until the decrease in energy per step reached 0.01 kcal mol^{-1}. Although HMs with small eigenvalues correspond to the soft degrees of freedom of the molecule, structural changes in these modes can lead to small deviations of the bonded geometry of the molecule. The SHAKE routine[40] was employed at each HM relaxation step in order to remove at least small deviations of the bond lengths from the optimal bond lengths. Depending on the subset of HMs, the CPU time (SGI, R8000) for a complete energy minimization in the HMs varied between 1 and 10 s. In the following the term HM relaxation is used to indicate an energy minimization in HMs. Energy minimization or full minimization refers to optimization in all atomic

coordinates including all pairwise intrareceptor interactions.

Results

ENERGY MINIMIZATION AND HARMONIC MODES OF DNA STRUCTURE

Four different ligand-free energy minimized DNA model structures (A_{free}–D_{free}) were generated using variations in the potential function describing the atomic interactions. Although the variations in force field parameters were rather small (see Methods section), the structural differences between the final energy minimized structures were significant. In particular, the variation of the dielectric constant (or scaling of charges) had a large effect on the final minimum energy structure. A difference in charge scaling between 0.37 and 0.33 resulted in structures with minor grooves slightly larger or smaller, respectively, than the minor groove of standard B-DNA. In addition, the final structure obtained with a charge scaling of 0.33 appears to be more compact than the B-DNA start structure. A possible explanation for this effect is that the backbone repulsion due to the negative charge of the phosphate groups tends to widen the minor groove. Below a certain charge scaling (around 0.35 in the present case, see Methods section) the attractive van der Waals interaction (between the two strands) out balances the backbone repulsion and in turn results in a smaller minor groove. Consistent with this hypothesis, the van der Waals interaction was found to be more favorable in the structure B_{free} compared to A_{free} (data not shown). An increase of the force constants for the improper torsion angle terms and the valence angle terms (case *C* and *D*) led to structures that were closer to B-DNA and appeared to be less sensitive to the choice of the dielectric constant (or charge scaling). The all atom root mean square deviation (RMSD) of the energy minimized structures from the standard B-DNA start structure ranged from 1.4 to 1.7 Å. The structural differences of energy minimized DNA structures in the presence of the minor groove ligand using the four force field variants (A_{bound}–D_{bound}) were smaller than the ligand-free minimized structures. These structures served as reference structures for comparison with ligand-bound forms relaxed in HMs (see below). For structures A_{bound}–D_{bound} the minor groove width was largely determined by the dimensions of the minor groove ligand. The minor groove width was smaller compared to the B-DNA start structures. The average RMSD between the energy minimized ligand-bound forms was < 0.5 Å and 1 Å compared to the corresponding segment in the X-ray structure.[26] The RMSD of A_{free}–D_{free} from the corresponding ligand-bound structures (A_{bound}–D_{bound}) were 0.95, 1.06, 1.05, and 0.76 Å, respectively. Structures A_{free} and C_{free}, as well as A_{bound} and C_{bound}, are shown as stereoplots in Figure 1.

Because each model structure (A_{free}–D_{free}) consists of 276 atoms, there are $3 \times 276 - 6 = 822$ vibrational modes with nonzero eigenvalues. Most of these HMs have large eigenvalues corresponding to structural deformations that disrupt the stereochemical geometry of the molecule and lead to atomic overlap. HMs with small eigenvalues correspond to internal motions of the DNA to some degree compatible with the stereochemical geometry. As an example to illustrate possible motions of the DNA associated with a set of HMs, the conformational change upon deforming structure D_{free} in the four modes of lowest eigenvalues is shown in Figure 2. Conformational changes associated with these HMs correspond to global structural changes of the DNA molecule (e.g., overall bending and major and minor groove opening and closing motions). The study of these collective modes and comparison with more detailed simulation methods such as molecular dynamics is of considerable interest for an understanding of the flexibility of DNA.[19,20] However, a systematic analysis and interpretation of the HMs of DNA and the force field dependence is not the focus of the current study.

CONFORMATIONAL DIFFERENCE BETWEEN LIGAND-BOUND AND -FREE DNA IN TERMS OF HARMONIC MODES

The eigenvectors of the second derivative matrix of the potential function for the receptor form a complete orthogonal basis to describe any conformation of the N atoms forming the molecule. The conformational difference between an energy minimized free DNA model and a model in the ligand-bound form can be described in terms of HM contributions. Because the HMs are orthogonal each contribution is given by the scalar product of the eigenvector (HM) and the Cartesian conformational difference between the bound and free receptor form.

Similar to, for example, an optical spectrum, all eigenvector contributions to the conformational

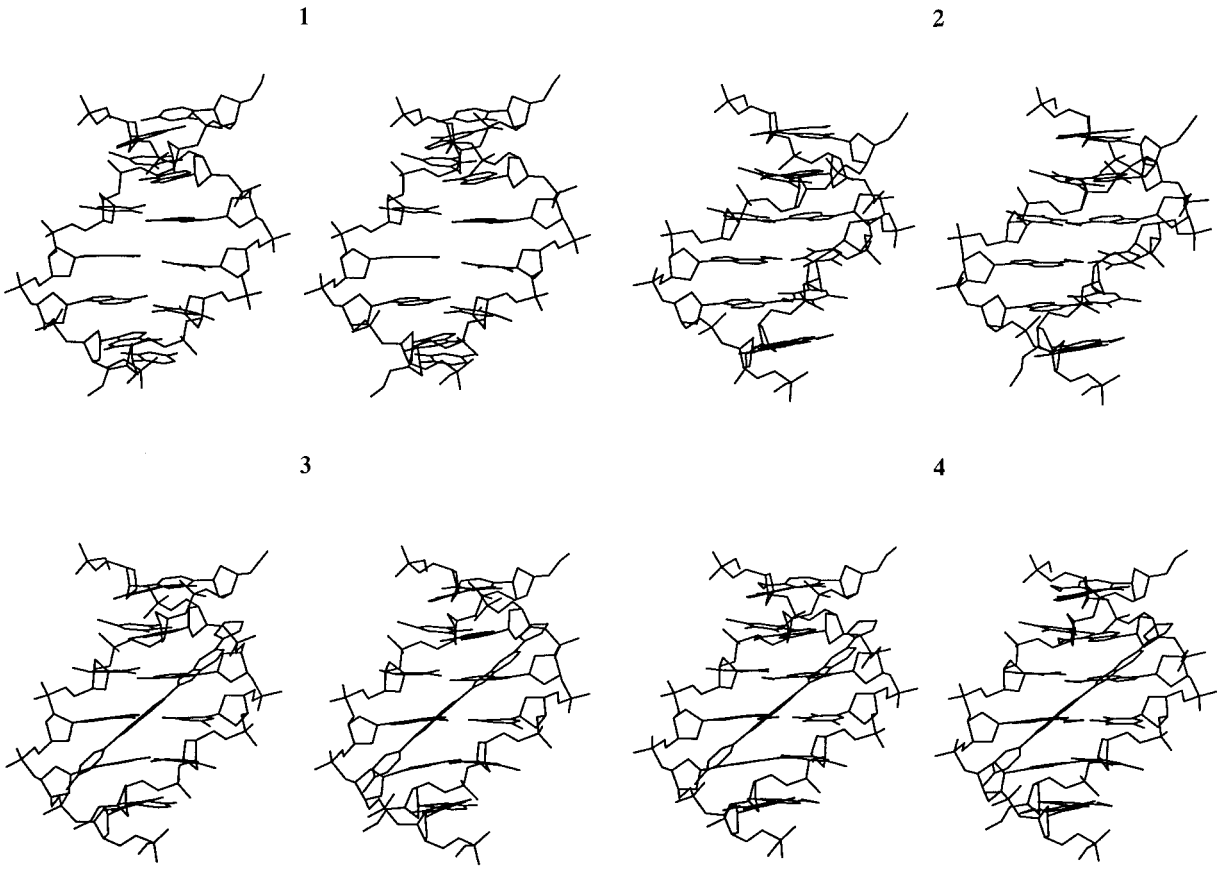

FIGURE 1. Energy minimized B-DNA structures A_{free}, C_{free} and A_{bound}, C_{bound}. For force field modifications *A* and *C* the resulting energy minimized structures A_{free}/C_{free} (energy minimized without ligand, upper panels 1 and 2) and A_{bound}/C_{bound} (minimized in the presence of the Hoechst 33258 analogue[26] in the DNA minor groove, lower panels 3 and 4), respectively, are shown as stereo plots. The structures are shown from approximately the same viewpoint looking into the DNA minor groove.

difference between the receptor in the complexed and free form could be termed a spectrum of the conformational difference in the HMs. This conformational difference spectrum for each of the four models, $A_{bound}-D_{bound}/A_{free}-D_{free}$, respectively, is shown in Figure 3. Most of the conformational difference between the receptor bound and free form can be described by a few harmonic modes (with low eigenvalues). There is a clear clustering in the spectrum in the range of HMs with small eigenvalues for all four models.

One can also use various sets of HMs to best approximate the conformational difference between bound and free receptor forms (Fig. 4). Depending on the model, inclusion of between five and 40 modes is necessary to reduce the RSMD between the complexed and free receptor conformations from initially ~ 1 Å down to ~ 0.2 Å. These results served as a guideline for the choice of HMs used as variables in the docking simulations.

RELAXATION OF DNA MODEL STRUCTURES IN HARMONIC MODES UPON LIGAND DOCKING

In a simple computational docking experiment the putative ligand is placed at various positions relative to the receptor molecule and a contact energy is calculated to assess the position and orientation as a possible binding site. In the current study, similar to most docking methods that assume a rigid receptor molecule, only a sterical (van der Waals) interaction term in the form of a Lennard–Jones function was used to evaluate a given ligand binding position. For a computational ligand–receptor docking approach it is desirable to recognize the ligand position identified in the ex-

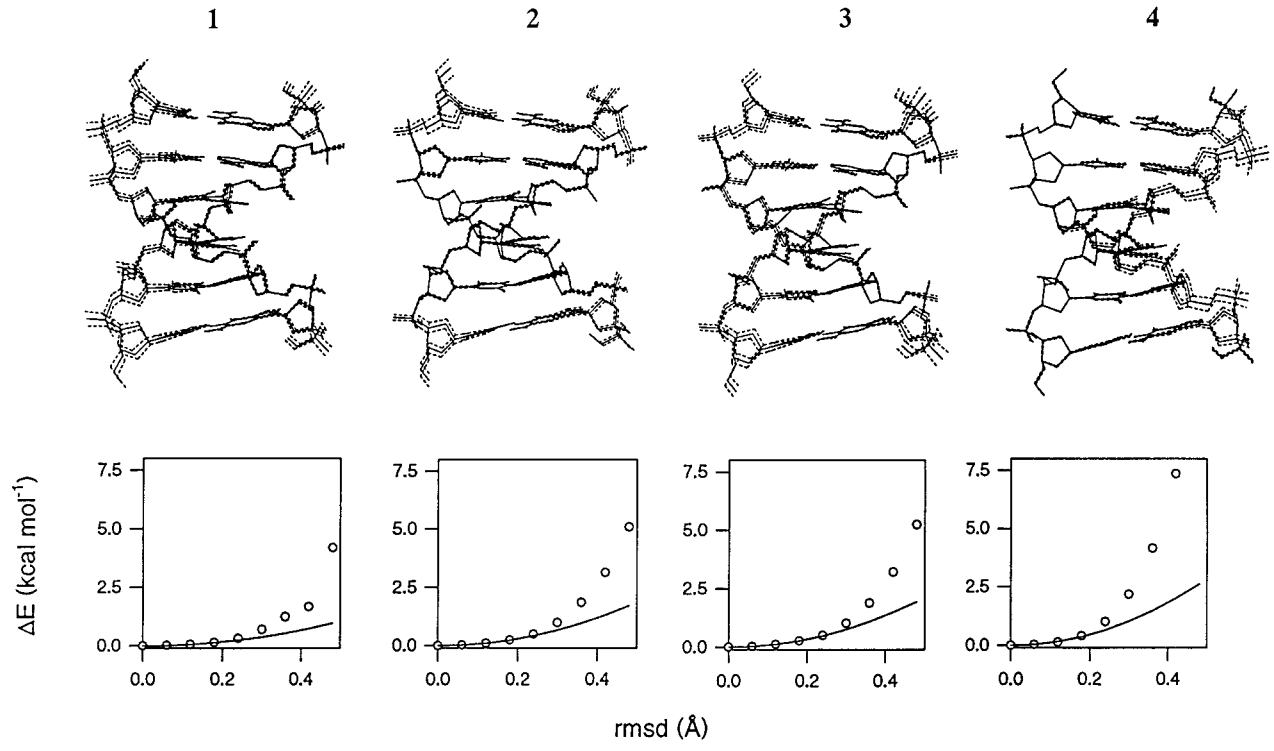

FIGURE 2. Deformation of the minimum energy structure D_{free} in the direction of the four HMs with the smallest eigenvalues. (1 –4): Stick plots of two structures (dashed lines) deformed by ± 0.12 Å (RMSD from D_{free}), respectively, in the direction of HMs 1 –4. Structure D_{free} is superimposed in bold. HM 1 is the mode with the lowest eigenvalue. The eigenvalue of each HM increases from 1 to 4. The stick plots are oriented such that the minor groove is on the left of each structure and the major groove is on the right. The motion associated with HM 1 consists of an overall bending and minor groove closing and opening deformation. Modes 2 and 3 correspond to overall bending motions (largest deformations at the ends of the structures) and some minor groove motion (HM 3); HM 4 consists largely of a major groove opening and closing deformation (little conformational change at the minor groove). The corresponding change in internal receptor energy upon deformation in the HM direction is plotted in the panel below each structure [(○) summing over all receptor atom pairs using force field model *D*]. For comparison, the change in receptor energy as suggested by the harmonic analysis is also plotted [(—) 0.5 × eigenvalue × (RMSD)^2].

perimental (X-ray) structure as a favorable binding site. As a test case, the minor groove binding ligand Hoechst 33258 analogue was placed into the minor groove of the DNA fragment in the conformation and position found in the corresponding X-ray structure of the complex.[26] In the structures A_{free} and D_{free} the Lennard–Jones interaction energy between the DNA and ligand was negative, indicating a possible ligand binding site already (Table I). For the model structures B_{free} and C_{free}, however, serious sterical overlap between the ligand and DNA resulted in a very large Lennard–Jones energy (Table I). Relaxation of the receptor structures in the HMs resulted in ligand–DNA complexes with significantly improved packing interactions between the receptor and ligand (Table I, Fig. 5). In models *B* and *C* the ligand binding site would have been missed by a method using a rigid receptor due to the unfavorable Lennard–Jones interaction. Depending on the number of HMs, the HM relaxation approach indicates a favorable ligand binding site. For force field models *A* and *D* the binding positions might be considered only as low score binding sites. Relaxation in the subset of HMs transforms these sites into much more favorable binding positions (Table I). The HM relaxed receptor structures can be compared to receptor–ligand complexes energy minimized using all atom Cartesian variables (A_{bound}–D_{bound}, minimized starting from B-DNA + ligand). The RMSD of the HM relaxed receptor structures (A_{HM}–D_{HM}) relative to these fully energy minimized bound forms was smaller in all cases than the RMSD of the corresponding free forms (vs. A_{bound}–D_{bound}). However, the RMSD did not reach the lowest possible level suggested

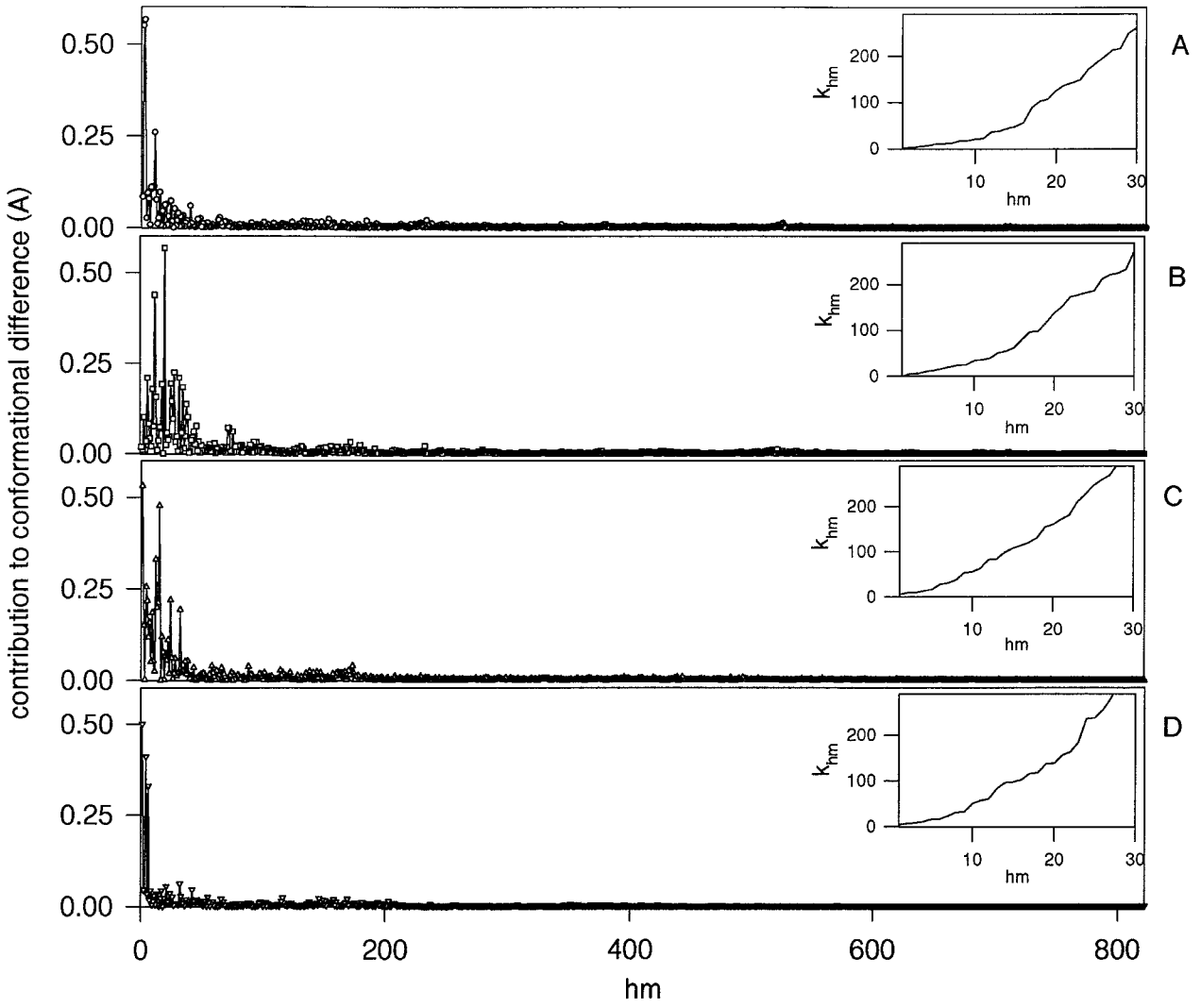

FIGURE 3. Harmonic mode components (spectrum) of the difference between the receptor in the ligand-bound form and corresponding free form. $(A-D)$ The conformational differences between DNA structures $A_{bound}-D_{bound}$ and $A_{free}-D_{free}$, respectively, in terms of HMs of the ligand-free conformation, are plotted vs. harmonic mode number (hm). The HMs are ordered according to corresponding eigenvalues (HM with the lowest eigenvalue has the number 1). The eigenvalues for the first 30 eigenvectors are displayed for each model in the insets to each panel $\{k_{hm}$ [kcal mol^{-1} Å (rmsd)$^{-2}$]$\}$.

from the 'spectrum' of the conformational difference between the free and bound forms (Figs. 3, 4). For models A and D, including more than 27 collective modes started to decrease the similarity of the HM relaxed structures relative to A_{bound} and D_{bound}, respectively. Small deformations in HMs corresponding to small eigenvalues of the receptor structure are to some degree compatible with the stereochemical geometry of the molecule. However, small changes in bond lengths and bond angles caused by the HM relaxation can lead to high intramolecular energies of the receptor. The structures $A_{HM}-D_{HM}$ in complex with the ligand can only be considered as preselected candidates for an analysis using a more accurate description of the receptor.

HM RELAXED RECEPTOR STRUCTURES AS STARTING STRUCTURES FOR ENERGY MINIMIZATION IN PRESENCE OF LIGAND

The receptor structures $A_{HM}-D_{HM}$ obtained as described in the previous paragraph were used as starting structures for energy minimization. Energy minimization was performed in Cartesian space including all atoms of the receptor (and all

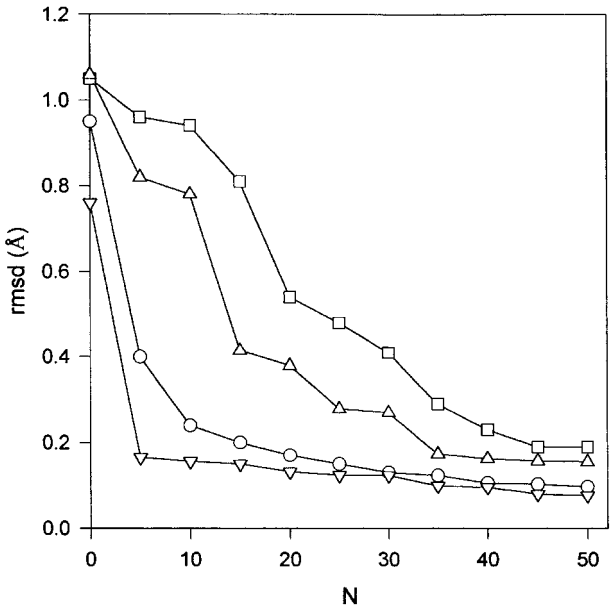

FIGURE 4. The root mean square deviation (Å) of a best possible approximation of the difference between ligand-free and -bound DNA forms is plotted as a function of the number (N) of HMs (with small eigenvalues) included in the approximation: (○) model A, (□) model B, (△) model C, (▽) model D.

pairwise intra-DNA interactions) in the presence of a fixed ligand at the position and orientation found in the X-ray structure. The number of minimization steps to reach an energy 5, 1.25, and 0.25 kcal mol^{-1}, respectively, above the energy found for the fully energy minimized DNA ligand complex (A_{bound}–D_{bound}) was recorded and compared to energy minimization using regular B-DNA or A_{free}–D_{free} start structures in the presence of the ligand. For all models, energy minimization from the start structures A_{HM}–D_{HM} converged significantly faster (by a factor of 2–3, see Table II) to receptor structures close to the fully minimized bound forms than starting from either B-DNA or A_{free}–D_{free}, respectively. Starting energy minimization from B_{free} or C_{free} led to energy minima for the ligand-bound receptor forms that were different from the corresponding structures B_{bound} and C_{bound}. (These were obtained using B-DNA as the start structure; see above.) However, the number of energy minimization steps to reach these minima was again larger than the corresponding number of energy minimization steps to reach B_{bound} or C_{bound} starting from B_{HM} or C_{HM}, respectively.

Most likely, the faster convergence starting from the HM relaxed receptor forms was due to the smaller RMSD relative to the corresponding fully minimized ligand-bound receptor structures compared to the minimized free or regular B-DNA receptor conformations (Table I, Fig. 5). The result demonstrates that structures A_{HM}–D_{HM} are reasonable start structures for refinement of the minor groove ligand–DNA complex using energy minimization converging in every case faster to low energy structures than starting from regular B-DNA or corresponding minimized ligand-free DNA structures.

DNA RECEPTOR HM RELAXATION IN PRESENCE OF MISPLACED LIGANDS

Most ligand–receptor docking approaches represent the receptor as a rigid structure. A given ligand binding pocket on a rigid receptor can accommodate only a limited number of ligands, ligand orientations, and ligand conformations. The limits are given by the fixed shape of the rigid receptor structure. In contrast, the current study treated the DNA molecule as an elastic body allowing deformations of the molecule in response to a bound ligand. This might lower the selectivity of the docking approach to discriminate between placements and orientations of the ligand different from the optimal ligand position in a receptor binding pocket (in the following termed misplacements). In order to test the selectivity of the approach for ligand misplacement, three reasonable alternative positions and orientations of the ligand relative to the DNA molecule (for force field variant A) were generated. These correspond to positions expected to result in favorable ligand–receptor interactions with a large contact area between the ligand and DNA: placement as observed in the X-ray structure,[26] placement in the DNA major groove, ligand rotated by 180° around its long axis in the minor groove, and like the previous one but position shifted along the minor groove. It needs to be emphasized that these alternative ligand positions form only a small subset of possible binding modes. Energy minimization with respect to translational and orientational degrees of freedom of the (rigid) ligand relative to the DNA was used to obtain structures with favorable DNA–ligand Lennard–Jones interaction (see Table III). Subsequently, the receptor structure was relaxed with respect to the 27 HMs with the lowest eigenvalues. As expected, minimization of the receptor structure with the misplaced ligands resulted in improved interactions between the receptor and ligand. However, the energy was in all

TABLE I.
Docking Minor Groove Binding Ligand to DNA Model Structures.

N_e	E_{TOT}	E_{LJ}	E_{PEN}	$rmsd_{start}$	$rmsd_{target}$
		A_{free}			
0	−46.9	−46.9	0.0	0.0	0.93
9	−85.3	−88.5	3.1	0.66	0.49
18	−87.4	−90.6	3.2	0.68	0.39
27	−91.1	−94.8	3.7	0.72	0.43
36	−95.3	−98.4	3.1	0.67	0.50
45	−96.3	−99.1	2.8	0.64	0.55
		B_{free}			
0	22751.0	22751.0	0.0	0.0	1.05
9	4076.0	3425.0	651.0	2.18	2.30
18	508.9	250.8	258.1	1.50	1.36
27	106.1	8.0	98.1	1.20	0.96
36	10.4	−53.3	63.7	1.10	0.83
45	−16.1	−69.3	53.2	1.04	0.71
		C_{free}			
0	132956.0	132956.0	0.0	0.0	1.00
9	1068.2	897.3	170.9	1.59	1.55
18	15.7	−38.3	54.0	1.0	0.61
27	−33.3	−68.0	34.7	0.96	0.56
36	−54.1	−81.6	26.5	0.94	0.55
45	−59.9	−84.0	24.1	0.91	0.56
		D_{free}			
0	−34.9	−34.9	0.0	0.0	0.76
9	−83.5	−86.7	3.2	0.61	0.32
18	−85.5	−89.0	3.5	0.63	0.35
27	−88.6	−92.1	3.5	0.62	0.38
36	−91.4	−94.3	2.9	0.57	0.44
45	−93.8	−96.2	2.4	0.57	0.48

N_e corresponds to the number of harmonic modes (with small eigenvalues) included in the HM relaxation of the receptor in the presence of ligand at the position found in the X-ray structure. HM relaxation was performed using the four model structures $A_{free}-D_{free}$. E_{TOT} is the total energy (kcal mol^{-1}) including Lennard–Jones interaction (E_{LJ}) between receptor atoms and ligand atoms and the penalty energy (E_{PEN}) given by the deformations in the HMs. $RMSD_{start}$ and $RMSD_{target}$ correspond to the root mean square deviation (Å) of the relaxed DNA structure from the starting structures ($A_{free}-D_{free}$) and fully energy minimized ligand–DNA receptor structures ($A_{bound}-D_{bound}$), respectively.

cases significantly higher than the energy of the HM relaxed structure with the ligand placed at the position and orientation found in the X-ray structure. At least for the three reasonable alternative ligand placements considered here, the approach still identified a position close to the X-ray structure as the most favorable ligand placement. Interestingly, the improvement in van der Waals interaction upon HM relaxation was also largest for the 'correct' placement of the minor groove ligand. It is important to note that the energy criterion used here to evaluate a possible binding site is only of qualitative value. It excludes, for example, any electrostatic and solvent effects.

Discussion

One limitation of most current computational ligand–receptor docking approaches is the assumption of a rigid receptor structure. Using all atom Cartesian variables and pairwise intrarecep-

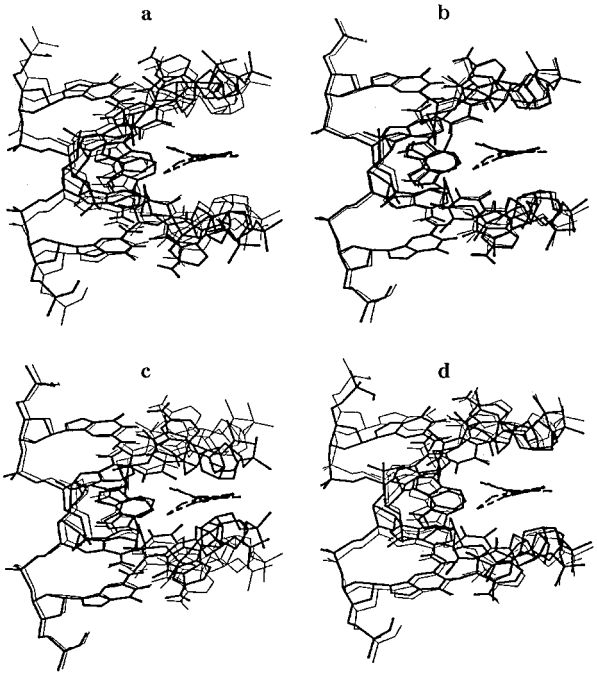

FIGURE 5. Comparison of HM relaxed DNA structures and energy minimized ligand-free and ligand-bound structures. (a) Superposition of A_{free} (bold line) on A_{bound} (thin lane); (b) superposition of A_{HM} (bold line, relaxed in 27 HMs) on A_{bound} (thin lane); (c) superposition of B_{free} (bold line) on B_{bound} (thin lane); (d) superposition of B_{HM} (bold line, relaxed in 45 HMs) on B_{bound} (thin lane). The minor groove ligand is shown using dashed lines. All structures are viewed along the minor groove.

tor interactions to account for receptor flexibility during the docking procedure is computationally very demanding, in particular when studying a large number of possible ligands as docking candidates. In the present investigation a subset of HMs corresponding to eigenvectors of the Hessian matrix were used as variables to approximately describe DNA–receptor flexibility. The DNA and minor groove ligand system was chosen because it is of moderate size compared to most protein–ligand complexes, allowing a relatively rapid calculation of the HMs. Given the DNA structure at an energy minimum, the modes with small eigenvalues correspond to the soft degrees of freedom (i.e., with a small energy change upon conformational deformation in these modes). Comparison of known X-ray and NMR structures revealed that, depending on the sequence and/or solution or crystallization conditions, DNA structures can vary in minor or major groove width and overall curvature.[41–43] Specifically, the binding of a ligand to the minor groove of DNA can result in a narrowing of the groove or widening of the minor groove in case of two ligands binding in a side by side mode.[26–31,36] To some degree these conformational variations of the DNA were captured by the HMs with small eigenvalues calculated for the present test system representing global conformational changes (i.e., overall stretching and bending and minor and ma-

TABLE II.
Energy Minimization of HM Relaxed DNA Ligand Complexes.

Model	Start Structure	+5	+1.25	+0.25
A	B-DNA	255	430	610
	A_{free}	110	210	410
	A_{HM}	28	91	220
B	B-DNA	250	420	605
	B_{free}	140	270	510
	B_{HM}	85	164	320
C	B-DNA	250	470	850
	C_{free}	950	1500	> 2000
	C_{HM}	77	153	279
D	B-DNA	280	510	880
	D_{free}	120	225	485
	D_{HM}	30	110	220

Energy minimization of minor groove / DNA complexes was performed starting from different start structures (column 2, B-DNA, $A_{free}-D_{free}$, $A_{HM}-D_{HM}$) and using four different force field variants (column 1). Columns 3–5 indicate the number of conjugated energy minimization steps in Cartesian atomic coordinates to reach an energy of 5, 1.25, and 0.25 kcal mol^{-1}, respectively, above the energy of the fully minimized complex ($A_{bound}-D_{bound}$).

TABLE III.
Ligand Misplacement.

Position	Initial Energy	Final Energy	Difference
1	−74.6	−87.5 / −89.3	−12.9 / −14.7
2	−56.5	−63.2 / −64.3	−6.7 / −7.8
3	−48.9	−52.6 / −53.2	−3.7 / −4.3
4	−55.1	−63.0 / −64.2	−7.9 / −8.1

The first column indicates the positions of the minor groove ligand relative to DNA (A_{free}); position 1 corresponds to a position close to the X-ray structure; 2–4 correspond to ''misplacements'' (see text for details). the initial energy (column 2) corresponds to E_{LJ} (see Table I) of the ligand and DNA after energy optimization of the orientational and translational degrees of freedom of the ligand. The final energy (column 3) indicates the energy ($E_{LJ}+E_{PEN}$) / Lennard–Jones energy (E_{LJ}), respectively, after HM relaxation (27 HMs). The last column is the improvement in interaction energy upon HM relaxation ($E_{LJ}+E_{PEN}/E_{LJ}$). All energies are in kilocalories per mole (kcal mol^{-1}).

jor groove opening and closing motions). For all force field variants only a relatively small number of HMs with small eigenvalues was necessary to model most of the conformational difference between a complex and free DNA receptor structure. This is a decisive prerequisite for the application of the HM relaxation approach to account for receptor flexibility in ligand docking simulations. The RMSD between the complexed and free form for the present test system was relatively small (~ 1 Å). It is likely that, for larger differences in particular, if potential energy barriers like torsion angle flips are involved, HM variables are not suitable to account for the conformational difference between free and bound receptor structures.

The HM relaxation protocol as devised in the present study led to a significantly improved sterical fit between the minor groove ligand and DNA in a simple docking attempt. Without HM relaxation the docking site that corresponded to the experimentally found ligand position could have been overlooked (model *B* and *C*) or counted only as a weak binding site (model *A* and *D*). In addition, the HM relaxation protocol resulted in structures with a significantly reduced RMSD relative to the bound receptor structure. Misplacements of the ligand were still identified on the basis of a reduced van der Waals interaction compared to the placement observed in the X-ray structure.

A possible component of the sequence-specific binding of minor groove ligands to DNA could be the sequence dependence of DNA flexibility that might, in part, also be accounted for using HMs. This was not tested; one reason is the very approximate nature of the current force field. A number of theoretical studies focused on the influence of ligands on DNA flexibility.[44–48] Ligand flexibility, solvent effects, and electrostatic interactions that were neglected in the present study may also be important components responsible for the sequence specificity of minor groove binding ligands.[49,50] It needs to be stressed that the focus of the current study is not to offer new insights on the specificity or affinity of minor groove ligands to DNA but to present a possible approach to improve simple ligand–receptor docking methods that usually employ rigid ligand and receptor models to account only for a reasonable sterical fit between the ligand and receptor. During the HM relaxation the HMs are treated as independent variables with a very simple penalty function for deviations in the HMs from the minimum energy receptor structure. This function allows for some conformational freedom in the HMs within the limits given by the penalty function. It underestimates the internal energy change of the receptor for deformations in the selected HMs, mostly because the HM relaxed structures show some perturbations of the optimal bonded geometry. Energy minimization is required to remove these perturbations in bond length, bond angles, and so forth, and to generate a more accurate DNA–ligand complex model structure. Normal mode calculations in terms of torsion angle variables may reduce these perturbations.[16,18,20] In the present case, a small number of energy minimization steps was sufficient to reduce the internal energy of the HM relaxed DNA structures in complex with the ligand. The HM relaxation procedure can be considered as a first filter to preselect possible ligand–receptor structures for further refinement. One might argue that because energy minimization including all pairwise intrareceptor interactions is required to evaluate a possible HM relaxed complex, one could use this energy minimization procedure in the first place. However, in a realistic computational docking experiment with several hundred ligand candidates and possible binding sites, the speed of the first preselection step might be more important than the precision of the preselected structures. Full energy minimization of the receptor structure for each case requires depending on the size of the receptor at least ~ 1000 s (for each putative ligand and ligand position) compared to a few seconds of CPU time using the HM relaxation approach. It needs to be emphasized that the HM relaxation approach in its present form is still computationally much slower than rigid receptor–ligand docking approaches (i.e., such as the program DOCK[1,2]). Also, the receptor ensemble docking method (see Introduction) by Knegtel et al.[14] is faster than the present approach. However, in contrast to the present approach the method by Knegtel et al.[14] depends on the availability of an ensemble of structures representing the receptor and its conformational flexibility. Due to the lower speed of the HM relaxation approach in the current form compared to the above-mentioned methods, it may be limited to cases where the ligand binding site is approximately known and the number of possible ligands is manageable.

A shortcoming of the HM relaxation approach is that it requires an energy minimization of the free receptor structure down to small residual gradients in order to precalculate the HMs. Energy minimization and diagonalization of the second derivative matrix of the potential energy function

can both become computationally difficult and very demanding for larger receptor structures. There are, however, efficient component synthesis, reduced basis set, and iterative methods available to extract subsets of eigenvectors from large matrices.[51–54] The HMs are all calculated prior to the docking simulation and can be used for every docking attempt using various possible ligands and ligand starting positions.

Another problem indirectly affecting the present approach is that, due to force field limitations, energy minimization can lead to conformational changes that can increase the RMSD between free and ligand-bound receptor structures. Comparison of crystal structures of unliganded DNA and DNA in complex with a minor groove ligand shows that the ligand induced conformational changes are in the same order as those caused by energy minimization of the uncomplexed DNA due to, for example, the limited accuracy of the force field and the neglect of the explicit solvent. In the current study, for computational reasons a relatively small DNA fragment was used as a test system to represent the receptor structure; this may be another reason for the deviation of the energy minimized structure from the regular B-DNA structure. It should be emphasized that it is possible to extend the approach to larger DNA fragments. The approach would benefit from a more realistic force field that reproduces the experimentally observed structure as an energy minimum. However, for the present test case and all force field variants that resulted in different energy minimized structures, it was still possible to assign most of the conformational difference between complexed and free DNA structures to a small set of eigenvectors.

For many receptor–ligand systems the ligand induced conformational changes in the receptor are small enough to justify the assumption of a rigid receptor during ligand docking. Besides the receptor targets for which the 3-dimensional (3-D) structure was determined experimentally, there are many interesting protein and nucleic acid receptors for which only model built structures can be generated. In case of proteins these model built structures are often based on homology to a protein with a known 3-D structure. Due to the same force field limitations discussed above and approximations inherent to the model building procedure, it is unlikely that these homology built structures contain (rigid) cavities that exactly fit a possible ligand candidate. Some degree of receptor flexibility is desirable. Here the HM relaxation approach might also be useful compared to docking procedures that use a rigid receptor structure.

Ligand binding can lead in some cases to a small global rearrangement of the protein main chain combined with side chain motions that involve torsion angle flips that cannot be modeled using a few HMs. In this case a mixed approach might be possible in which the side chain rearrangement is considered explicitly (using, e.g., side chain torsion angles as variables and main chain flexibility using HMs).

Although the focus of the present study was to approximately account for receptor flexibility, the approach could in principle also be used to approximately account for ligand flexibility during docking. However, in many cases conformational changes of the ligand upon docking are much larger than changes in the receptor. In particular, if these changes involve, for example, significant changes in some torsion angles including crossing of barriers (torsion angle flips), the possibility to assign most of the conformational changes to a few HMs breaks down.

The HM relaxation approach in combination with energy minimization of preselected docking candidates might be useful as an extension of rigid docking methods to account for small ligand induced conformational changes of the receptor at modest additional computational cost. However, the assignment of ligand induced conformational changes to a small number of HMs depends on the energy minimum structure and force field. The approach would benefit from force field improvements to yield energy minimized structures closer to the experimentally determined conformation. Applications to other ligand receptor systems are also necessary to further test the approach.

References

1. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. J Mol Biol 1982, 161, 269.
2. Kuntz, I. D.; Meng, E. C.; Shoichet, B. K. Acc Chem Res 1994, 27, 117.
3. Lybrand, T. P. Curr Opin Struct Biol 1995, 5, 224.
4. Blaney, J. M.; Dixon, J. S. Perspect Drug Discov Design 1993, 1, 301.
5. Böhm, H.-J.; Klebe, G. Angew Chem 1996, 108, 2750.
6. Sobolev, V.; Wade, R. C.; Vriend, G.; Edelman, M. Proteins 1996, 25, 120.
7. Böhm, H.-J. J Comput Aided Mol Design 1992, 6, 61.
8. Eisen, M. B.; Wiley, D. C.; Karplus, M.; Hubbard, R. E. Proteins 1994, 19, 199.
9. Welch, W.; Ruppert, J.; Jain, A. N. Chem Biol 1996, 3, 449.

10. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. J Mol Biol 1996, 261, 470.
11. Leach, A. R. J Mol Biol 1994, 235, 345.
12. Zacharias, M.; Luty, B. A.; Davis, M. E.; McCammon, J. A. J Mol Biol 1994, 238, 455.
13. Luty, B. A.; Wasserman, Z. R.; Stouten, P. F. W.; Hodge, C. N.; Zacharias, M.; McCammon, J. A. J Comput Chem 1995, 16, 454.
14. Knegtel, R. M. A.; Kuntz, I. D.; Oshiro, C. M. J Mol Biol 1997, 266, 424.
15. Levy, R. M.; Karplus, M. Biopolymers 1979, 18, 2465.
16. Go, N.; Noguti, T.; Nishikawa, T. Proc Natl Acad Sci USA 1983, 80, 3696.
17. Brooks, B. R.; Karplus, M. Proc Natl Acad Sci USA 1983, 80, 6571.
18. Levitt, M.; Sander, C.; Stern, P. S. J Mol Biol 1983, 181, 423.
19. Tidor, B.; Irikura, K. K.; Brooks, B. R.; Karplus, M. J Biomol Struct Dyn 1983, 1, 231.
20. Duong, T. H.; Zakrzewska, K. J Comput Chem 1997, 18, 796.
21. Karplus, M.; Kushick, J. N. Macromolecules 1981, 14, 325.
22. Levy, R. M.; Perahia, D.; Karplus, M. Proc Natl Acad Sci USA 1982, 79, 1346.
23. Kitao, A.; Hirata, F.; Go, N. J Chem Phys 1991, 158, 447.
24. Hayward, S.; Kitao, A.; Hirata, F.; Go, N. J Mol Biol 1993, 234, 1207.
25. Amadei, A.; Linssen, B. M.; Berendsen, H. J. C. Proteins 1993, 17, 412.
26. Wood, A. A.; Nunn, C. M.; Czarny, A.; Boykin, D. W.; Neidle, S. Nucl Acids Res 1995, 23, 3678.
27. Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. J Mol Biol 1985, 183, 553.
28. Pjura, P.; Grzeskowiak, K.; Dickerson, R. E. J Mol Biol 1987, 197, 257.
29. Teng, M. K.; Usman, N.; Frederick, C. A.; Wang, A. H. Nucl Acids Res 1988, 16, 2671.
30. Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingam, M. Nature Struct Biol 1994, 1, 169.
31. Chen, X.; Ramakrishnan, B.; Sundaralingam, M. Nature Struct Biol 1995, 2, 733.
32. Kopka, M. L.; Goodsell, D. S.; Han, G. W.; Chiu, T. K.; Lown, J. W.; Dickerson, R. E. Structure 1997, 5, 1033.
33. Jenkins, T. C.; Lane, A. N.; Neidle, S.; Brown, D. G. Eur J Biochem 1993, 213, 1175.
34. Blasko, A.; Browne, K. A.; Bruice, T. C. J Am Chem Soc 1994, 116, 3726.
35. Pelaez, R.; de Clairac, L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. J Am Chem Soc 1997, 119, 7909.
36. Neidle, S. Biopolymers 1997, 44, 105.
37. QUANTA 96 Molecular Modeling Package; Molecular Simulations Inc., University of York: York, U.K., 1997.
38. van Gunsteren, W. F.; Berendsen, H. J. C. Groningen Molecular Simulation (GROMOS) Library Manual; BIOMOS B.V.: Groningen, The Netherlands, 1987.
39. Press, D. W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. Numerical Recipes; Cambridge University Press: Cambridge, U.K., 1989.
40. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. J Comput Phys 1977, 23, 327.
41. Drew, H. R.; Dickerson, R. E. J Mol Biol 1981, 151, 535.
42. Goodsell, D. S.; Kaczor−Grzeskowiak, M.; Dickerson, R. E. J Mol Biol 1994, 239, 79.
43. Berman, H. M. Biopolymers 1997, 44, 23.
44. Rao, S. N.; Kollman, P. A. Proc Natl Acad Sci USA 1987, 84, 5735.
45. Boehnke, K.; Nonella, M.; Schulten, K. Biochemistry 1991, 30, 5465.
46. Herzyk, P.; Neidle, S.; Goodfellow, J. M. J Biomol Struct Dyn 1992, 10, 97.
47. Chen, Y. Z.; Prohofsky, E. W. Biopolymers 1994, 35, 657.
48. Duong, T. H.; Zakrzewska, K. J Biomol Struct Dyn 1997, 14, 691.
49. Singh, S. B.; Wemmer, D. E.; Kollman, P. A. Proc Natl Acad Sci USA 1994, 91, 7673.
50. Zakrzewska, K.; Madami, A.; Lavery, R. Chem Phys 1996, 204, 263.
51. Harrison, R. W. Biopolymers 1984, 23, 2943.
52. Hao, M. H.; Harvey, S. C. Biopolymers 1992, 32, 1393.
53. Brooks, B. R.; Janezic, D.; Karplus, M. J Comput Chem 1995, 16, 1522.
54. Mitin, A. V. J Comput Chem 1994, 15, 747.